



AFRL-RH-WP-TR-2008-0074

**Multi-Aircraft Video - Human/Automation
Target Recognition Studies:
Video Display Size
in Unaided Target Acquisition
Involving Multiple Videos**

**Sarah E. Plantz
Lamar Warfield
Thomas R. Carretta
Airam Gonzalez-Garcia
Michael J. Patzek**

**Warfighter Interface Division
Supervisory Control Interfaces Branch
Wright-Patterson AFB OH 45433-7511**

**April 2008
Final Report for July 2007 to November 2007**

**Approved for public release;
Distribution is unlimited.**

**Air Force Research Laboratory
Human Effectiveness Directorate
Warfighter Interface Division
Supervisory Control Interfaces Branch
Wright-Patterson AFB, OH 45433**

NOTICE AND SIGNATURE PAGE

Using Government drawings, specifications, or other data included in this document for any purpose other than Government procurement does not in any way obligate the U.S. Government. The fact that the Government formulated or supplied the drawings, specifications or other data does not license the holder or any other person or corporation; or convey any rights or permission to manufacture, use, or sell any patented invention that may relate to them.

This report was cleared for public release by the Air Force Research Laboratory Wright Site Public Affairs Office and is available to the general public, including foreign nationals. Copies may be obtained from the Defense Technical Information Center (DTIC) (<http://www.dtic.mil>).

TECHNICAL REVIEW AND APPROVAL

AFRL-RH-WP-TR-2008-0074

THIS TECHNICAL REPORT HAS BEEN REVIEWED AND IS APPROVED FOR PUBLICATION.

FOR THE DIRECTOR

//signed//

Sarah E. Plantz
Computer Scientist
Supervisory Control Interfaces Branch

//signed//

Daniel G. Goddard
Chief, Warfighter Interface Division
Human Effectiveness Directorate

This report is published in the interest of scientific and technical information exchange, and its publication does not constitute the Government's approval or disapproval of its ideas or findings.

REPORT DOCUMENTATION PAGE				<i>Form Approved</i> OMB No. 0704-0188		
Public reporting burden for this collection of information is estimated to average 1 hour per response, including the time for reviewing instructions, searching data sources, gathering and maintaining the data needed, and completing and reviewing the collection of information. Send comments regarding this burden estimate or any other aspect of this collection of information, including suggestions for reducing this burden to Washington Headquarters Service, Directorate for Information Operations and Reports, 1215 Jefferson Davis Highway, Suite 1204, Arlington, VA 22202-4302, and to the Office of Management and Budget, Paperwork Reduction Project (0704-0188) Washington, DC 20503.						
PLEASE DO NOT RETURN YOUR FORM TO THE ABOVE ADDRESS.						
1. REPORT DATE (DD-MM-YYYY) 21-04-2008		2. REPORT TYPE Interim		3. DATES COVERED (From - To) July 2007 – November 2007		
4. TITLE AND SUBTITLE Multi-Aircraft Video – Human/Automation Target Recognition Studies: Video Display Size in Unaided Target Acquisition Involving Multiple Videos				5a. CONTRACT NUMBE		
				5b. GRANT NUMBER NA		
				5c. PROGRAM ELEMENT NUMBER 62202F		
				5d. PROJECT NUMBER 7184		
6. AUTHOR(S) Sarah E. Plantz Lamar Warfield Thomas R. Carretta Airam Gonzalez-Garcia Michael J. Patzek				5e. TASK NUMBER 09		
				5f. WORK UNIT NUMBER 17		
7. Air Force. PERFORMING ORGANIZATION NAME(S) AND ADDRESS(ES) Air Force Material Command Air Force Research Laboratory Human Effectiveness Directorate Warfighter Interface Division Wright-Patterson AFB, OH 45433				8. PERFORMING ORGANIZATION REPORT NUMBER		
9. SPONSORING/MONITORING AGENCY NAME(S) AND ADDRESS(ES) Air Force Material Command Air Force Research Laboratory Human Effectiveness Directorate Warfighter Interface Division Wright-Patterson AFB OH 45433				10. SPONSOR/MONITOR'S ACRONYM(S) AFRL/RHCI		
				11. SPONSORING/MONITORING AGENCY REPORT NUMBER AFRL-RH-WP-TR-2008-0074		
12. DISTRIBUTION AVAILABILITY STATEMENT Approved for public release; Distribution is unlimited.						
13. SUPPLEMENTARY NOTES 88 ABW PA Cleared 05/23/08, WPAFB-08-3390.						
14. ABSTRACT Prior research involving target detection when monitoring multiple video sources led to concerns about increased operator workload and degraded target detection performance as the number of video sources and visual search requirements increased. A study was conducted to examine the utility of smaller video displays to reduce visual scan and search demands when multiple video sources are monitored. Sixteen participants performed a target detection task involving one small video, one large video, four small videos, and four large videos. Results indicated no significant differences in target detection performance and workload for small vs. large video sources. However, in the four-video condition, participants made significantly more false alarms when viewing the small videos. Further, when asked to indicate their preference, participants preferred the large videos, but the difference was statistically significant in only the single video condition. Implications for video display designs involving monitoring multiple video sources are discussed.						
15. SUBJECT TERMS Unaided target detection, unmanned air systems, video display size, workload						
16. SECURITY CLASSIFICATION OF: Unclassified			17. LIMITATION OF ABSTRACT UNL		18. NUMBER OF PAGES 45	
19a. NAME OF RESPONSIBLE PERSON Sarah E. Plantz			19b. TELEPHONE NUMBER (Include area code)			
a. REPORT U	b. ABSTRACT U	c. THIS PAGE U				

CONTENTS

PREFACE	1
INTRODUCTION	1
METHOD	2
Participants	2
Measures	2
MAV Videos	3
Equipment	4
Procedures	5
Analyses	9
RESULTS	9
Target Identification Accuracy	9
Confidence Ratings	12
Workload	13
Video Quality and Interpretability	15
Target Attributes, Task Characteristics, and Target Detection Accuracy	18
DISCUSSION	21
REFERENCES	24
APPENDIX	25

APPENDICES

A	Study Questionnaires	25
B	Candidate Video Display Formats for DVR Implementation	32

FIGURES

1	Computer monitors used to display pictures of the targets and MAV videos	4
2	Ground-based pictures of targets	6
3	Example of four large (5 by 6 ¾", 640 x 480 resolution) videos, four small (2 ½ by 3 5/16", 320 x 240 resolution) videos, one large (5 by 6 ¾", 640 x 480 resolution) video, and one small (2 ½ by 3 5/16", 320 x 240 resolution) video	7
4	Example of video placement for a participant	8
5	Hit accuracy (%) by number of videos and video display size	10
6	Number of false alarms by number of videos and video display size ...	12
7	Mean confidence rating by number of videos and video display size	13
8	Overall workload by number of videos and video display size	14
9	Image interpretability rating by number of videos and video display size ..	16
10	Number of participants that preferred small or large video display size for one and four video display conditions	18
A-1	NASA Task Load Index computer application used	31
B-1	Four live only video condition (Baseline)	33
B-2	Four live videos with embedded DVR video condition	34
B-3	Four live videos with a single auxiliary DVR capability condition	35
B-4	Four live videos with four auxiliary DVR capabilities condition	36
B-5	Four live videos (top) with four auxiliary DVRs capabilities condition ...	37

TABLES

1	Target Identification Accuracy: Target Identification Percent by Number of Videos and Video Display Size Combination	10
2	Target Identification Accuracy: Number of False Alarms by Number of Videos and Video Display Size Combination	11
3	Confidence Rating by Number of Videos and Video Display Size Combination	13
4	Overall Workload by Number of Videos and Display Size Combination ..	14
5	Image Characteristics: Descriptive Statistics for Image Quality, Clarity, Contrast, Resolution, and Interpretability Ratings	15
6	Image Characteristics: Image Interpretability Rating by Number of Videos and Video Display Size Combination	16
7	Participants' Preferences based on Video Display Size: Image Interpretability Rating by Number of Videos and Video Display Size Combination	17
8	Descriptive Statistics for Target Characteristics (Size and Time on Screen) and Target Detection Accuracy Grouped by Task Characteristics	19
9	Correlations among Target Attributes, Task Characteristics, and Target Detection Accuracy	21

PREFACE

This report describes activities performed in support of the Air Force Research Laboratory Warfighter Interface Division, System Control Interfaces Branch (AFRL/RHCI) Interfaces for Small Unmanned Systems, Work Unit 71840917. The authors thank Mr. Greg Feitshans, Mr. Allen Rowe, and the entire Vigilant Spirit Team for providing the technical support that allowed this study to occur.

THIS PAGE INTENTIONALLY LEFT BLANK

INTRODUCTION

As more unmanned aerial systems are incorporated into everyday military operations and as their roles become more demanding, efforts are underway to advance operator interface technology to improve human performance, system capability, and overall mission effectiveness. Some employment concepts leverage unmanned systems to act as force multipliers, where a single operator may be expected to simultaneously monitor and exert executive control over several unmanned systems (Barbato, 2000; Walker, 2005). The System Control Interfaces Branch of the Air Force Research Laboratory is exploring multi-UAV, single-operator concepts for conducting reconnaissance, surveillance, and target acquisition (RSTA) missions. The multi-UAV research includes investigating advanced operator-vehicle interface technology and human performance associated with both mission and sensor management.

Warfield, Carretta, Patzek, and Gonzalez-Garcia (2007) explored the relationships between the number of micro air vehicle (MAV) videos monitored and target detection performance. The study provided baseline data to better understand human performance with single and multiple videos, without decision support or performance-enhancing technology (e.g., automatic target cueing). Results indicated that participants could detect most of the targets and the number of targets detected was not significantly affected by the number of videos monitored. However, as the number of videos monitored increased from one to two to four, participants were more likely to make false alarms (i.e., designate objects as targets when they were not targets). Confidence in target detection decisions generally declined as the number of video displays increased.

Warfield et al (2007) suggested several technology-based methods for improving target detection performance: synthetic overlays, target cueing, image enhancement, and digital video recording (DVR)/playback. Subsequent research on the DVR/playback methods uncovered potential problems. When multiple video presentation with DVR capability was prototyped and demonstrated, concerns arose that adding the DVR functionality might further increase visual search and control task requirements and detract attention from the live videos, possibly increasing workload and adversely affecting target detection performance. One particular design issue and possible trade-off was the number and size of displays that were required to present the real-time and recorded videos. Given this, the research team decided to examine the utility of

smaller video displays in order to reduce visual scan and search demands when multiple video sources were to be used.

The current study's objective was to compare target acquisition performance using different MAV video display sizes (i.e., 5" by 6 3/4" with 640 x 480 resolution versus 2 1/2" by 3 5/16" with 320 x 240 resolution) for both single and multiple video presentations. Potential advantages of smaller video displays include less area to scan, a sensation of less video jitter, and more flexibility in display configuration and design. A potential drawback of smaller video displays is that the targets of interest will be smaller on the screen and be represented by fewer display pixels, which may adversely affect target acquisition performance. The findings and lessons learned could provide valuable insights into the size and configuration of the sensor display for performing target acquisition and contribute to shaping employment concepts and technology requirements for future unmanned aerial systems.

METHOD

Participants

Participants were 16 civilian and military full-time employees stationed at Wright-Patterson AFB, OH. The sample consisted of 15 men (93.8%) and one woman (6.3%). Participants ranged in age from 24 to 51 years with a mean age of 32.8 years. All participants reported being in good to excellent health and having visual acuity corrected to 20/20, normal color vision, and normal peripheral vision. Most participants indicated they had previous experience with simulators (63%) and video games (56%). Participation was voluntary and participants could withdraw from the study at any time without penalty. No compensation was offered in exchange for participation in this study.

Measures

The questionnaires used in this study are described below and provided in Appendix A. These were a Demographic Data/Background Questionnaire, Confidence Ratings, the National Aeronautics and Space Administration Task Load Index (NASA-TLX: Hart & Staveland, 1988), and a Post-Test Questionnaire.

Demographic data/Background Questionnaire. This questionnaire was used to collect information in order to characterize the participants in terms of prior experience and

demographic characteristics and assist in interpretation of participants' performance on the target detection task. Items elicited information about participants' sex, age, general health, wellbeing, previous experience with simulator-type environments, previous experience with video games, and whether they had vision correctable to 20/20 acuity and normal peripheral and color vision.

Objective measures of target acquisition performance. Two measures of target acquisition performance were collected: number of hits and number of false alarms. Number of hits was converted to a percentage as there were different video arrangements and number of targets in the one and four video configurations.

Confidence Ratings. Whenever participants detected a target, they were instructed to verbally state the type of target (shelter, SUV, truck, and van) and the level of confidence in their target detection decision. Confidence ratings were made on a five-point Likert rating scale (1 - not at all confident, 2 - slightly confident, 3 - moderately confident, 4 - fairly confident, 5 – very confident).

NASA-TLX. The NASA TLX (Hart & Staveland, 1988) is a subjective workload assessment tool. A multidimensional weighting procedure is used to derive an overall workload score based on weighted averages of ratings on 6 subscales: Mental, Physical, Temporal, Effort, Performance, and Frustration. See Figure A-1.

Post-Test Questionnaire. This questionnaire elicited information regarding participants' assessment of the video imagery used in the study. Participants rated the video imagery in terms of their perceived ability to accomplish the task on a five-point scale (1 - poor, 2 - fair, 3 - good, 4 - very good, 5 – excellent). Participants also were given the opportunity to provide comments regarding video quality and other factors that affected their ability to detect targets.

MAV Videos

The video used in this study was recorded from a forward-looking color camera mounted in the nose of a micro air vehicle (MAV) at a 45 degree depression. The camera had a resolution of 640 X 480 lines, 30 degree field of view, and a 2.4 GHz downlink (wireless data link) for video with a 900 MHz 2-way modem. The video was streamed at 30 frames per second. The MAV flew at approximately 175 feet altitude above the ground with an air speed of approximately 22 knots. Several videos of about 15-20 minutes in length were edited to create 12 five-minute clips for use as test material and five one-minute clips for pre-test training materials.

Several buildings, roads, and vehicles were dispersed over the setting. The targets used in the experiment were ground objects that were of interest during the Cooperative Operations in Urban Terrain (COUNTER) program's flight test and captured in the video recordings.

Equipment

Displays: Two side by side 24-inch widescreen LCD monitors were used to display still images of the targets and the videos. The still images of the target were provided to aid the participants during target acquisition. Both monitors had a resolution of 1920 x 1200 pixels. As illustrated in Figure 1, the still targets were displayed on the left monitor and the videos were displayed on the right monitor.



Figure 1. Computer monitors used to display pictures of the targets (left) and the MAV video(s) (right). A headset microphone was used to record voice to capture target identification and confidence ratings. Confidence rating values and meaning were displayed on a card under the right monitor.

Voice recording: As previously noted, whenever participants detected a target, they were instructed to verbally state the type of target and the level of confidence in their target detection decision. Participants' vocal responses were recorded using a Plantronics DSP 500 headset with a microphone.

Workload recording: A laptop positioned on a different table was used to run the computer-based NASA-Task Load Index application.

Procedures

Overview. The study was conducted in the AFRL/RHCI's Crew Systems Integration Laboratory (CSIL) at Wright-Patterson AFB, OH. The study required about two hours and was completed during a single testing session with a short break in the middle. Initial procedures included an overview of the study, informed consent, demographic data collection, familiarization with the equipment, and an explanation of the NASA-Task Load Index. Participants then completed a practice session to familiarize themselves with the equipment, procedures, and target acquisition task. Following each target acquisition trial (practice and test), participants completed the NASA-TLX.

The test trials were randomized across participants and occurred in a counterbalanced order that took into account video display size (large vs. small) and number of videos (1 vs. 4), and video arrangement (A vs. B). Post-study procedures included data collection regarding participant's comments on the target acquisition task and a study debriefing.

Initial procedures. Data collection began with completion of the demographic data questionnaire. Participants then completed a practice target acquisition session in order to familiarize them with the equipment and procedures.

Target acquisition. Participants were required to locate, designate, and identify targets that appeared in videos. Two side by side monitors were used. The right monitor displayed the videos while the left monitor displayed ground-based pictures and names of the targets to be found in the videos. Figure 2 shows the pictures of the targets used during the test trials. The task was to locate and correctly identify targets in the MAV videos. Participants were instructed that when they observed a target embedded in a video to use the mouse to place the cursor on the target, as close to the target as they could, and click the mouse. In addition, the participants were instructed to call out the name of the target with their confidence rating within two seconds after

they clicked on a target. Confidence ratings were on a five point scale (1- not at all confident, 2 – slightly confident, 3 – moderately confident, 4 – fairly confident, and 5 – very confident). The confidence rating scale and its values were displayed on a card under the right monitor throughout data collection as a reference. Objective performance measures for target identification accuracy were hits and false alarms. Subjective measures included confidence in target identification decisions and workload.



Figure 2. Ground-based pictures of targets.

The number of videos monitored simultaneously was varied to be either one or four. Video display size also was varied to be either 5 by 6 $\frac{3}{4}$ " with 640 x 480 resolution or 2 $\frac{1}{2}$ by 3 $\frac{5}{16}$ " with 320 x 240 resolution. This resulted in four number of videos by display size combinations: 1) one large (5 by 6 $\frac{3}{4}$ ", 640 x 480 resolution) video, 2) one small (2 $\frac{1}{2}$ by 3 $\frac{5}{16}$ ", 320 x 240 resolution) video, 3) four large (5 by 6 $\frac{3}{4}$ ", 640 x 480 pixels) videos, and 4)

four small (2 ½ by 3 5/16”, 320 x 240 resolution) videos. Examples of each of these are shown in Figure 3.



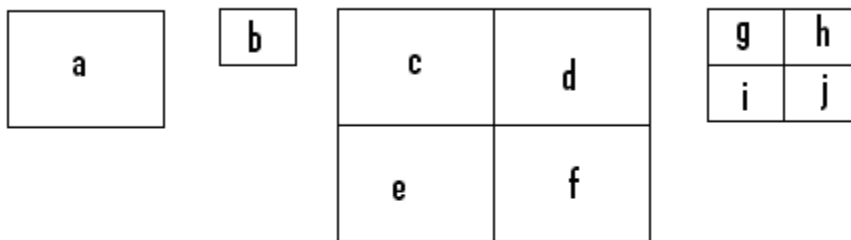
Figure 3. Example of four large (5 by 6 ¾”, 640 x 480 resolution) videos, four small (2 ½ by 3 5/16”, 320 x 240 pixels) videos, one large (5 by 6 ¾”, 640 x 480 resolution) video, and one small (2 ½ by 3 5/16”, 320 x 240 pixels) video. Each video or set of four videos was shown by itself in the middle of the right monitor with a black background.

The practice trials used a different target set and video footage than did the test trials. This was done so participants could become familiar with the target identification task procedures, but not with the test stimuli. The practice trials were each one-minute in length and occurred in the fixed order of one small video, one large video, four small videos, and four large videos. Following each target identification practice trial, participants completed the NASA-TLX.

During the test trials, each number of videos (1 or 4) by display size (large or small) combination occurred twice. There were 10 unique videos, 2 were used in the single video displays and 8 (2 combinations of 4 videos) were used in the four-video displays. Each of the videos was shown twice, once in the small (2 ½ by 3 5/16”, 320 x 240 resolution) format and

once in the large (5 by 6 ¾”, 640 x 480 resolution) format. The assignment of order for the four different display number/size combinations was randomized. The ordering of the videos was balanced across the participants; half of the participants viewed arrangement A first and the other half viewed arrangement B first. The viewing order was counterbalanced for number of videos (1 or 4), display size (large or small), and arrangement (A or B) (see Figure 4). Within the number of video conditions (1 vs. 4 videos), the same videos were used (i.e., videos used in the 1 video condition [videos a and b] were never used in the 4 video condition [videos c through j] and vice versa).

Trial 1 - Arrangement A



Trial 2 - Arrangement B

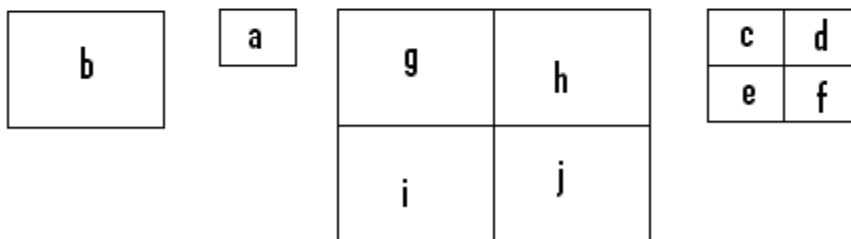


Figure 4. Example of video placement for a participant. Each video (a through j) was used twice, once in a large (5 by 6 ¾”, 640 x 480 resolution) video display and once in a small (2 ½ by 3 5/16”, 320 x 240 resolution) video display.

Each test trial was 5 minutes in length. Immediately following each video combination, participants were asked to rate their workload perception for that trial using the NASA-TLX. The procedure was repeated until all eight video presentations were completed. After viewing all of the videos, participants completed a questionnaire regarding image interpretability and quality.

Analyses

The purpose of the study was to compare the objective and subjective data on a target acquisition task for two video display sizes. The findings were intended to help guide the design and prototyping of advanced display concepts and subsequent studies, especially those requiring simultaneous monitoring of multiple videos. Although the study design crossed display size and number of video displays, the analyses focused on comparing performance within the number of videos conditions. That is, the focus was on comparing performance for one video (small vs. large display) and for four videos (small vs. large display). The decision not to examine the interaction of display size and number of videos on performance was due to the study design in that different sets of videos were used in the one and four video display conditions. As a result, the causes for any differences in pattern of performance between the one and four video conditions could be confounded by the different video sets used.

This was an exploratory study as we had no expectations as to which display size would be more effective based on results from prior studies. A .05 Type I error rate was used with a non-directional (two-tailed) hypothesis.

Analyses examined performance for display size (small vs. large) within each number of video displays condition (1 or 4). Related-samples t-tests were performed since participants were exposed to both display sizes. Objective measures of performance were hit accuracy (%) and number of false alarms. . Number of hits was converted to a percentage as there were different video arrangements and number of targets in the one and four video configurations. Subjective measures were overall workload, confidence in target detection decisions, and image interpretability ratings.

RESULTS

Target Identification Accuracy

Table 1 summarizes target identification performance by number and size of videos and Figure 5 provides an illustration. The mean target identification percent was 77.96% for the one video condition and 68.31% for the four video condition. Related samples t-tests indicated that within the number of videos conditions (1 or 4 videos), there was no significant difference in target identification percent for the small and large video sizes.

Table 1. Target Identification Accuracy: Target Identification Percent by Number of Videos and Video Display Size Combination

Condition	Mean	SD	Min.	Max.	SD _D	t (15)
<u>One Video</u>						
Small	78.12	9.63	60.00	90.00	8.64	0.14
Large	77.81	6.57	70.00	90.00		
<u>Four Videos</u>						
Small	68.31	10.47	44.83	82.76	10.53	0.00
Large	68.31	9.27	51.72	79.41		

Note. A 2-tailed related samples t-test was used to compare the mean difference in target detection percent for small vs. large displays within number of videos. SD_D is the standard deviation for the related samples t-test. Degrees of freedom (df) equals N pairs – 1 = 15. N = 16; * p < .05 (2-tailed)

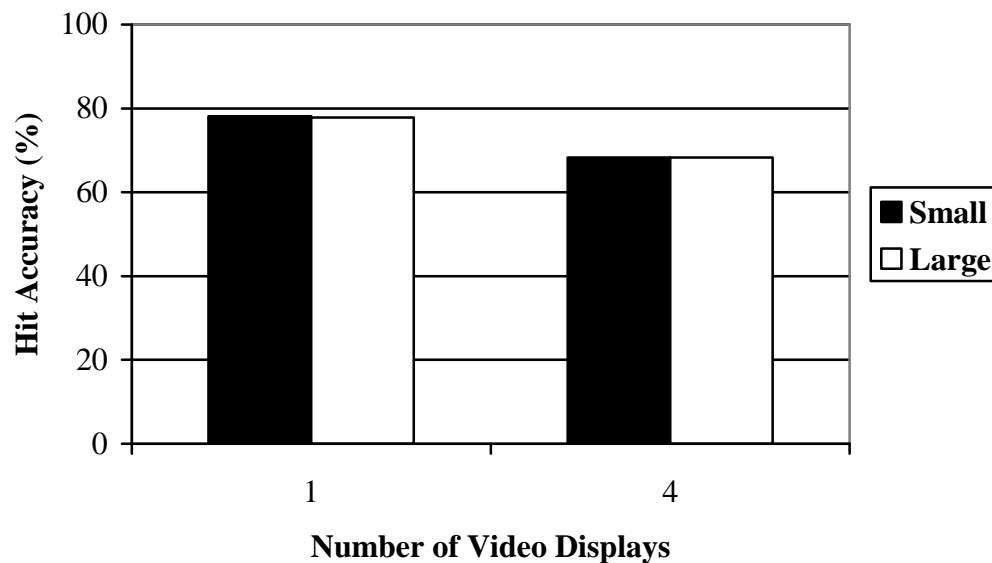


Figure 5. Hit accuracy (%) by number of videos and video display size.

Table 2 summarizes the number of false alarms made by number and size of videos and Figure 6 provides an illustration. The mean number of false alarms was 1.40 for the one video condition and increased to 2.56 for the four video condition. The related samples t-tests indicated that for the single video condition, there was no significant difference in the number of false alarms for the two display sizes ($t(1, 15) = -0.18$, ns). However, the number of false alarm was significantly greater for the small display compared to the large display when four videos were viewed ($t(15) = 2.58$, $p < .05$).

Table 2. Target Identification Accuracy: Number of False Alarms by Number of Videos and Video Display Size Combination

Condition	Mean	SD	Min.	Max.	SD _D	$t_{(15)}$
<u>One Video</u>						
Small	1.37	1.70	0	5	1.34	-0.18
Large	1.43	1.86	0	5		
<u>Four Videos</u>						
Small	3.25	2.62	0	10	2.12	2.58*
Large	1.87	1.40	0	4		

Note. A 2-tailed related samples t-test was used to compare the mean difference in number of false alarms for small vs. large displays within number of videos. SD_D is the standard deviation for the related samples t-test. Degrees of freedom (df) equals N pairs – 1 = 15. N = 16; * $p < .05$ (2-tailed)

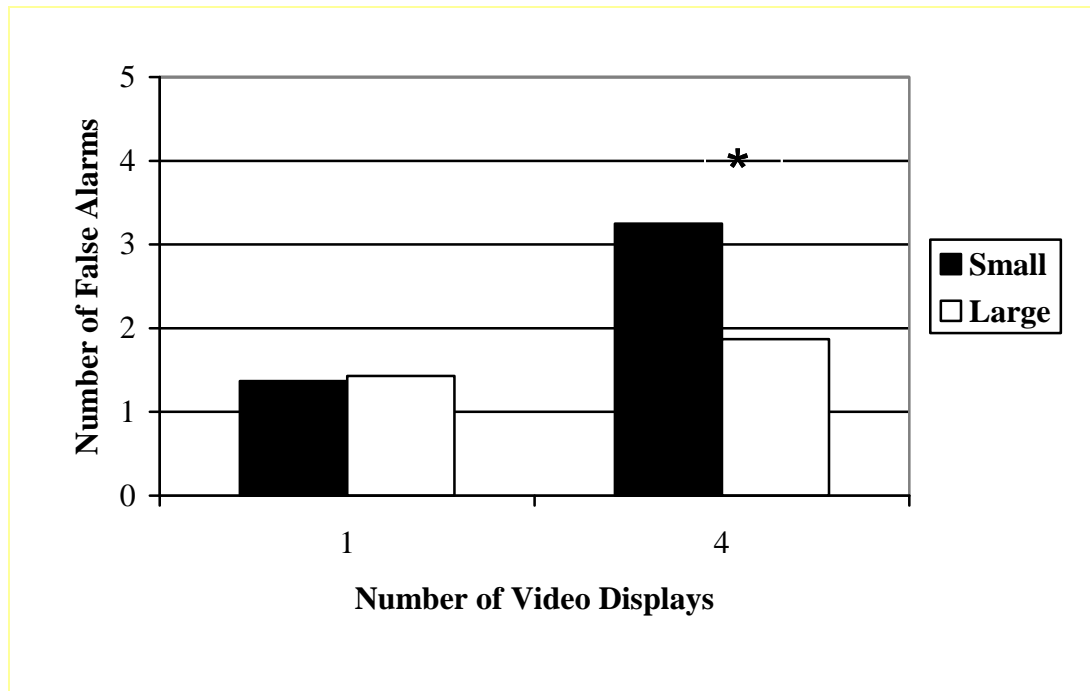


Figure 6. Number of false alarms by number of videos and video display size.

* $p < .05$

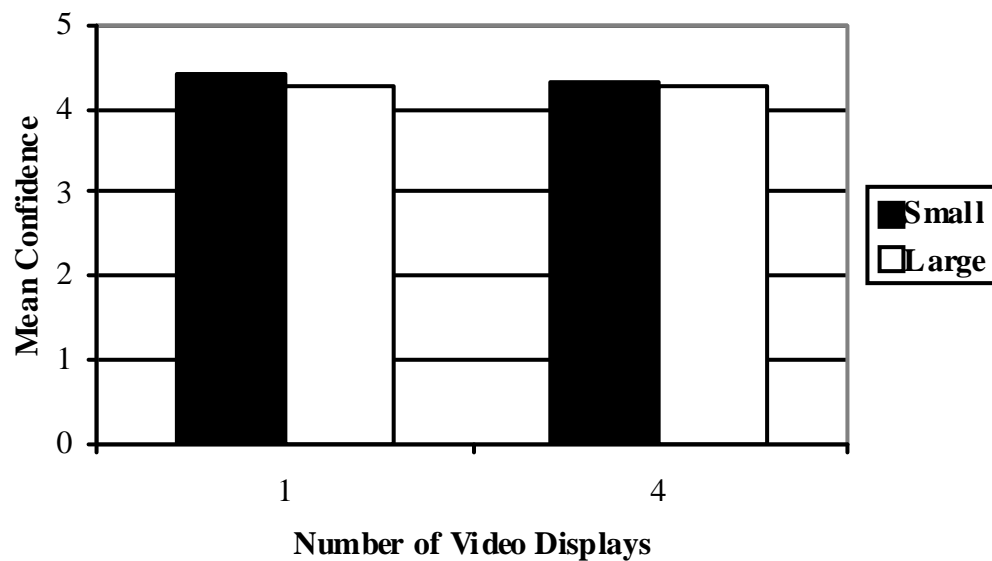
Confidence Ratings

A summary of participants' confidence ratings in their target identification decisions by number and size of video displays is provided in Table 3. Mean confidence ratings are illustrated in Figure 7. As shown in Table 3, there was little variability in confidence across the number of videos by display size combinations. The related samples t-tests indicated there was no significant difference in average confidence rating for the two display sizes within each of the number of video display conditions.

Table 3. Confidence Rating by Number of Videos and Video Display Size Combination

Condition	Mean	SD	Min.	Max.	SD _D	t (15)
<u>One Video</u>						
Small	4.40	0.43 [`]	3.47	5.00	0.32	0.89
Large	4.33	0.59	3.07	5.00		
<u>Four Videos</u>						
Small	4.25	0.53	3.06	5.00	0.29	-0.26
Large	4.27	0.53	2.93	5.00		

Note. A 2-tailed related samples t-test was used to compare the mean difference confidence ratings for small vs. large displays within number of videos. SD_D is the standard deviation for the related samples t-test. Degrees of freedom (df) equals N pairs – 1 = 15.

**Figure 7. Mean confidence rating by number of videos and video display size.**

Workload

Table 4 summarizes the average overall workload by number and size of video displays and Figure 8 provides an illustration. The mean overall workload was 43.36 for the one video

display condition and 54.77 for the four video display condition. The related samples t-tests indicated that there was no significant difference in average overall workload for the two display sizes within each of the number of video display conditions.

Table 4. Overall Workload by Number of Videos and Video Display Size Combination

Condition	Mean	SD	Min.	Max.	SD _D	t (15)
<u>One Video</u>						
Small	45.04	15.57	10.83	68.50	7.35	1.82
Large	41.68	15.62	10.67	69.50		
<u>Four Videos</u>						
Small	55.15	20.47	15.50	87.83	17.98	-1.84
Large	54.40	20.97	12.67	88.00		

Note. A 2-tailed related samples t-test was used to compare the mean difference in overall workload for small vs. large displays within number of videos. SD_D is the standard deviation for the related samples t-test. Degrees of freedom (df) equals N pairs – 1 = 15. N = 16; *p < .05 (2-tailed)

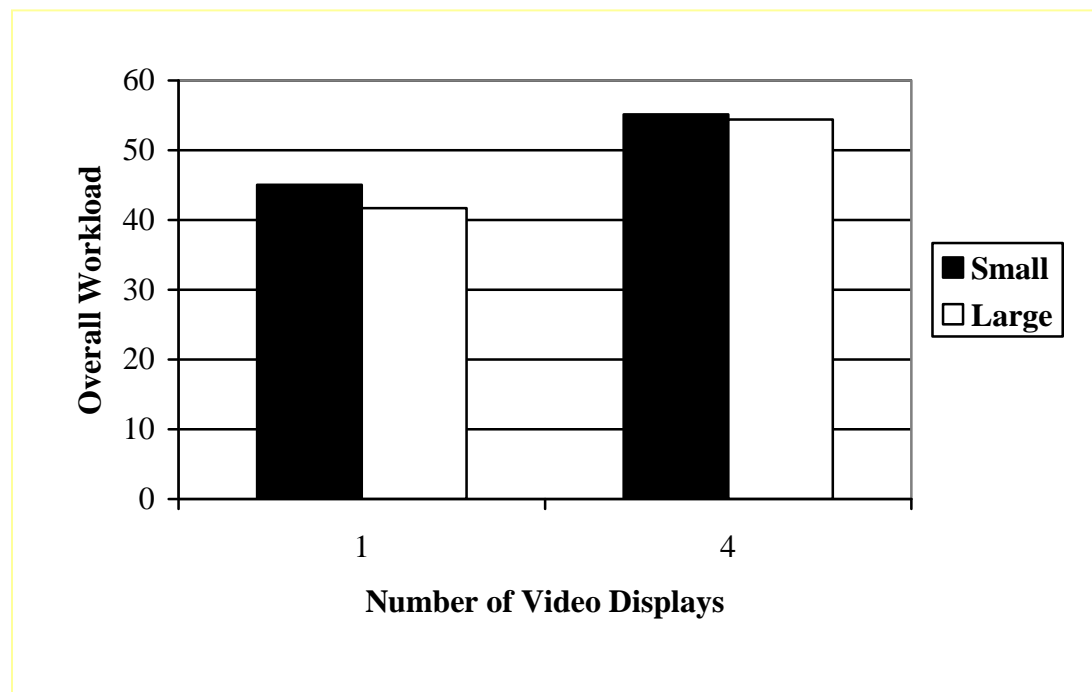


Figure 8. Overall workload by number of videos and video display size.

Video Quality and Interpretability

After completion of the target identification task, participants rated the quality, clarity, contrast, resolution, and interpretability of the video imagery used in the study. Video quality, clarity, contrast, and resolution were measured using a five-point scale (1 – poor, 2 – fair, 3 – good, 4 – very good, 5 – excellent). Interpretability was measured as a dichotomous variable (1 – yes, 0 – no). These data are summarized in Table 5. Mean ratings for quality, clarity, contrast, and resolution were all below the “3 – good” threshold. Eighty-eight percent of the participants rated the imagery as “interpretable.”

Table 5. Image Characteristics: Descriptive Statistics for Image Quality, Clarity, Contrast, Resolution, and Interpretability Ratings

Characteristic	Mean	SD	Min.	Max.
Quality	2.06	0.77	1	3
Clarity	2.38	0.80	1	3
Contrast	2.69	0.70	2	4
Resolution	2.63	0.80	1	4
Interpretability	0.88	0.34	0	1

N = 16

Participants also rated image quality/interpretability separately for the one small, one large, four small, and four large video display conditions. Ratings were on a five-point scale (1 – very poor, 2 – poor, 3 – fair, 4 – good, 5 – very good). Table 6 summarizes the mean image quality/interpretability ratings by number and size of video displays. Figure 9 provides an illustration. The related samples t-tests comparing the display sizes within the number of displays conditions indicated that image quality/interpretability was significantly lower for the small; display relative to the large display in the one video display condition. Although the direction of the difference was the same for the four display condition, the difference was not statistically significant. See Table 6 for a summary of these analyses.

Table 6. Image Characteristics: Image Quality/Interpretability Rating by Number of Videos and Video Display Size Combination

Condition	Mean	SD	Min.	Max.	SD _D	t (15)
<u>One Video</u>						
Small	3.50	0.73`	2	5	0.44	-8.47**
Large	4.44	0.62	3	5		
<u>Four Videos</u>						
Small	3.00	0.81	2	4	0.96	-1.81
Large	3.44	0.62	2	4		

Note. A 2-tailed related samples t-test was used to compare the mean difference in image interpretability ratings for small vs. large displays within number of videos. SD_D is the standard deviation for the related samples t-test. Degrees of freedom (df) equals N pairs – 1 = 15. N = 16; *p < .05; **p < .01 (2-tailed)

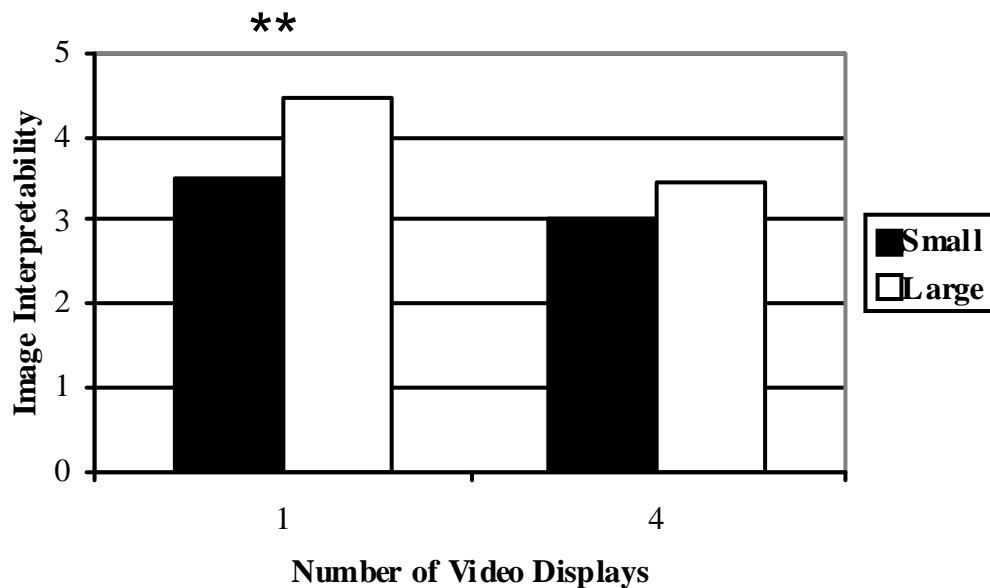


Figure 9. Image interpretability rating by number of videos and video display size.

**p < .01

An additional nonparametric analysis, the Wilcoxin matched pairs test, was used with the image quality/interpretability scale ratings to examine participants' video display size preference. The Wilcoxin matched pairs test is a nonparametric test used to examine distributional differences in performance for related groups. It calculates the difference between each set of pairs, ranks the differences (positive or negative), and analyzes that list of differences. If the two sums of ranks are very different, the probability level will be small. A significant difference indicates a preference for one display size over the other, but does not quantify the size of the difference.

Results for the Wilcoxin matched pairs test were consistent with those of the related samples t-tests reported in Table 6. Participants showed a strong preference for the large display format when one video was being viewed ($T = -3.63$, $p < .01$). Although the direction of the difference was the same for the four video display configuration, the difference between the small and large display formats was not significant. See Table 7 and Figure 10 for a summary.

Table 7. Participants' Preferences based on Video Display Size: Image Interpretability Rating by Number of Videos and Video Display Size Combination

Condition	Mean	SD	Negative Ranks	Positive Ranks	Tie Ranks	<i>T</i>
<u>One Video</u>						
Small	3.50	0.73	0	14	2	-3.63**
Large	4.44	0.62				
<u>Four Videos</u>						
Small	3.00	0.81	3	11	2	1.69
Large	3.44	0.62				

Notes. Positive ranks occurred when the large video display received a higher rating than did the small video display. Negative ranks occurred when the large video display received a lower rating than did the small video display. Tie ranks occurred when the large and small video displays received the same rating.

N = 16; * $p < .05$; ** $p < .01$

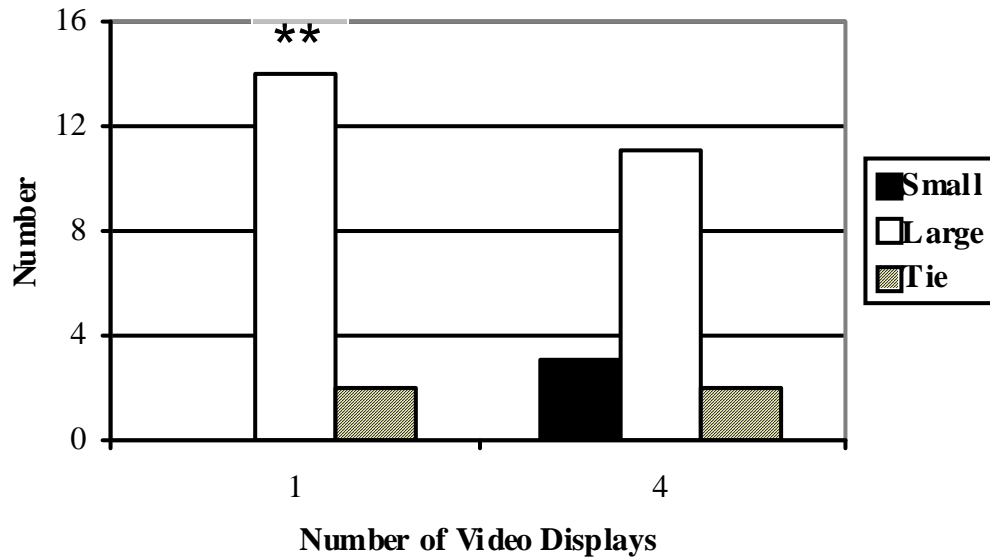


Figure 10. Number of participants that preferred small or large video display size for one and four video display conditions.

** $p < .01$

Target Attributes, Task Characteristics, and Target Detection Accuracy

Post-hoc analyses were performed to examine the relations between target attributes (target size, length of time the target was viewable) and task characteristics (number of video displays) and target detection probability in order to improve our understanding of the factors affecting target detection accuracy. There were 98 targets in all; two single videos with 10 targets each, one four-video condition with 14 targets and one four video condition with 15 targets. Each of these video arrangements was viewed twice, once in the small screen format and once in the large screen format ($2 \text{ display sizes} * (10 + 10 + 14 + 15 \text{ targets}) = 2 * 49 = 98 \text{ targets}$).

Examination of the descriptive statistics for the target attributes indicted substantial variability for target size and time on screen and for target detection accuracy for the 98 targets. See Table 8 for a summary. Mean time on screen for all 98 targets was 2.61 seconds, with a minimum of 0.20 seconds and a maximum of 9.92 seconds. Both the minimum and maximum time on screen targets were presented in the one video screen condition. Mean size for the 49

small targets was 1,393.56 pixels and ranged from 234.38 to 9,724.13 pixels¹. Mean target size for the 49 large targets was 5,224.82 pixels and ranged from 937.50 to 38,896.50 pixels. Mean hit rate for all 98 targets was 72.07% and ranged from 0.00% to 100.00%. A closer examination of the data revealed that one of the 98 targets was never detected and it occurred in the one small video presentation. Also, 13 of the 98 targets were detected 100.00% of the time (3 in the one small video condition, 7 in the one large video condition, 0 in the four small videos condition, and 3 in the four large videos condition).

Table 8: Descriptive Statistics for Target Characteristics (Size and Time on Screen) and Target Detection Accuracy Grouped by Task Characteristics

Condition	N	Time on Screen (seconds)			
		Mean	SD	Min	Max
<u>All Targets</u>	98	2.61	1.67	0.20	9.92
<u>One Video</u>					
Small	20	3.12	2.16	0.20	9.92
Large	20	3.12	2.16	0.20	9.92
<u>Four Videos</u>					
Small	29	2.25	1.15	0.30	5.10
Large	29	2.25	1.15	0.30	5.10
.....					
Condition	N	Target Size (pixels)			
		Mean	SD	Minimum	Maximum
<u>All Targets</u>	98	3,309.19	5,631.72	234.38	38,896.50
<u>One Video</u>					

¹ Target sizes in pixels are reported as non-integer values because they were calculated as average size taken by sampling each target several times as it was displayed.

Small	20	965.06	848.48	234.38	3,028.67
Large	20	3,004.08	2,189.75	937.50	10,706.67

Four Videos

Small	29	1,689.08	2,263.71	263.55	9,724.13
Large	29	6,756.36	9,054.85	1,054.20	38,896.50

.....

Hit Rate (%)					
Condition	N	Mean	SD	Minimum	Maximum
<u>All Targets</u>	98	72.07	27.84	0.00	100.00

One Video

Small	20	77.81	29.34	0.00	100.00
Large	20	77.86	29.92	6.25	100.00

Four Videos

Small	29	68.75	25.71	6.25	93.75
Large	29	67.45	27.41	6.25	100.00

Note. N = 98 targets

Table 9 shows the correlations among the target and display attribute variables and target detection accuracy. Both time on screen ($r = 0.525$, $p < .01$) and number of video displays ($r = -0.173$, $p < .05$) were correlated significantly with target detection accuracy (hit %), while target size ($r = -0.078$, ns) was not. A regression model that used time on screen, target size, and number of videos was significantly related to target hit percent ($R = 0.538$, $p < .01$). However, this model was not significantly different from one that used only time on screen ($r = 0.525$). These results are consistent with previous empirical studies regarding the probability of target detection in time-limited search (Wilson, Devitt, & Maurer, 2005). Wilson et al. demonstrated a strong non-linear mathematical relationship between time available for search and probability to detect a target. In their model, the probability to detect a target was nearly 0% when time

available to search was less than 0.8 seconds, increased steeply as time approached 3 seconds, then increased at a much lower rate.

Table 9. Correlations among Target Attributes, Task Characteristics, and Target Detection Accuracy

Variable	Time on Screen	Target Size	N Video Displays	Target Hit %
Time on Screen	1.000			
Target Size	-0.049	1.000		
N Video Displays	-0.249**	0.196*	1.000	
Target Hit %	0.525**	0.078	-0.173*	1.000

Notes. Target size was measured in pixels. Time on screen was measured in seconds. Target hit % reflected the percent of participants that detected the target.

N = 98 targets; * $p < .05$; ** $p < .01$

DISCUSSION

This study was initiated as a precursor to a larger study intended to evaluate the effectiveness of digital video recording (DVR) as a tool to aid the micro aerial vehicle operator in target detection. In the build up to testing two trial runs were conducted amongst members of the Interfaces for Small Unmanned Systems (ISUS) team in order to try to eliminate design options and determine the level of task difficulty. The primary task of the DVR study is to evaluate how the use of video playback affects an operator's performance in target detection while monitoring four separate live video streams. At the time of the current study, the proposed display designs for the DVR (video playback) study were constrained to displaying all videos (live and playback) across a single monitor. The use of a single monitor prohibited placing eight (four live and four playback) "large" windows each sized at 640 X 480 pixels. Eight windows are needed because some designs used a window for viewing each of four live videos and a corresponding window for each playback (DVR) video. As a result, it became necessary to

develop four “small” windows (i.e., 320 X 240 pixels each) in order to accommodate the different designs.

When performing these preliminary evaluations of the proposed designs (see Appendix B) the research team noticed that the live MAV video appeared less jumpy in the small format than those seen in the large format. This led us to speculate that people may be able to detect targets more easily on a small display than on the larger format. The working hypothesis was: given the target objects were large enough to be detected the small display format would require fewer foveal fixation points by the human eyes to cover than that of a larger area. Speed of detection, which some consider an important aspect of performance, especially in time-critical tasks was not measured in the current study. Future studies should include both detection accuracy and speed.

Many of the candidate targets used in the current study (vehicles, buildings) were fixed objects in the flight test range where the videos were recorded. In order to determine the smallest objects detectable in the MAV videos, the MAVs flew over four 4 ft. x 7 ft. modified Tri-Bar charts with varying levels of contrast: 1) white and black, 2) light gray and black, 3) mid-range gray and black, and 4) dark gray and black). The length of each strip was 12 inches. The width of the bars were as follows 1 inch, 2 inches, 4 inches, 6 inches, 8 inches, 12 inches, and 24 inches. The charts were laid end to end in an open field. The images of the four Tri-Bar charts were not included in the study because not enough usable video of them was captured. When viewing the images of the Tri-Bar charts on the large and small windows it was observed that none of the contrasting bars were visible in the video. Because the bars were not distinguishable, we limited our target set to vehicles and shelters.

Given the tendency for higher false alarms for the four display condition and that it was significantly higher for the four small displays versus the four large displays perhaps there are design methods that can be developed to mitigate overall false alarms reporting with multi-video presentation and, additionally, negate the false alarm differences between the display sizes. One concept is to add some type of confirmation method to the target detection and identification sequence. For example, once the operator suspects and indicates a target is in the video, a still image or portion of the video could automatically be displayed to allow the operator to further assess and verify the target is present. Accuracy and confidence may increase with such a

method. The downside, however, may be significant delays in reporting targets and possibly an increase in misses given the attention required for the confirmation process.

What implications do these results have for future employment concepts? If target detection accuracy, without making errors such as false alarms is crucial, the large display formats are preferable. This type of scenario would occur if the concept of employment was that the person doing the target detection task also was making the final decision as to whether or not to designate and attack a suspected target. If this is not the case, and the initial target detection decision will be reviewed by another source, then the higher false alarm rate observed for the small display format is not so important. The ability to use a smaller display size may provide greater flexibility to human factors engineers when developing display formats for systems with limited viewing areas.

REFERENCES

- Barbato, G. (2000). Uninhabited combat air vehicle controls and displays for suppression of enemy air defenses. *CSERIAC Gateway*. 11 (1), 1-4.
- Hart, S. G., & Staveland, L. E. (1988). Development of NASA TLX (task load index): Results of empirical and theoretical research in human mental workload. In P. A. Hancock & N. Meshkati (Eds.), *Human mental workload* (pp. 139-183). New York: Elsevier.
- Walker, B. (2005). *Virtual environment UAV swarm management using GPU calculated digital pheromones*. Unpublished doctoral dissertation, Iowa State University, Ames, IO.
- Warfield, L., Carretta, T. R., Patzek, M. J., & Gonzalez-Garcia, A. (2007). *Multi-aircraft video-human/automation target recognition studies: Unaided target acquisition involving multiple micro air vehicle (MAV) videos*, AFRL-HE-WP-TR-2007-0036. Wright-Patterson AFB, OH: Air Force Research Laboratory, Warfighter Interface Division.
- Wilson, D., Devitt, N., & Maurer, T. (2005). Search times and probability of detection in time-limited search. Infrared imaging Systems: Design, Analysis, Modeling, and Testing XVI, G. C. Holst (Ed.). *Proceedings of SPIE*, 5784, 224-231.

APPENDIX A
Study Questionnaires

Demographic Data Questionnaire

Participant ID: _____

1. Age: _____
2. Gender (circle one) Male Female
3. Describe your general health (circle one):

 Poor Fair Good Very Good Excellent
4. How would you assess your overall feeling of wellbeing this morning/afternoon (circle one)?

 Poor Fair Good Very Good Excellent
5. Do you have any practical experience working in a simulation type environment?

If yes explain:
6. Do you play any type of computer/video games? Yes No
 - a. If you answered “Yes,” what types do you play? (circle all that apply)

Action/Adventure _____ Role Playing _____
Other (specify) _____
 - b. Do the computer/video games you play require you to do visual search tasks (i.e., locate/identify objects or targets)? Yes No
7. Is your visual acuity correctable to 20/20? Yes No
8. Do you have any problems with your peripheral vision? Yes No
9. Are you color blind? Yes No
10. Are you aware you may withdraw from this study at any time? Yes No
11. Are you aware that your participation is strictly confidential? Yes No

Post-Test Interview Questions

1. How would you rate the quality of the video presented on this display device? (circle one)

Poor Fair Good Very Good Excellent

If your answer to #1 was “poor” or “fair,” what factors affected your rating?

2. How would you assess the clarity of the video imagery? (circle one)

Poor Fair Good Very Good Excellent

3. How would you assess the contrast of the video imagery? (circle one)

Poor Fair Good Very Good Excellent

4. How would you assess the resolution of the video imagery? (circle one)

Poor Fair Good Very Good Excellent

5. Did the display provide a sufficiently interpretable image? (circle one) Yes No

If your answer to #5 was “No,” what factors affected your rating?

6. Were you able to identify all predefined targets of interest in the video? (circle one)

Yes No

If no explain:

7. Please rate the video interpretability of the single screen small and large video displays and the four screen small and large video displays:

	<u>One Screen</u>		<u>Four Screens</u>	
	Small	Large	Small	Large
Very Good	<div></div>	<div></div>	<div></div>	<div></div>
Good	<div></div>	<div></div>	<div></div>	<div></div>
Fair	<div></div>	<div></div>	<div></div>	<div></div>
Poor	<div></div>	<div></div>	<div></div>	<div></div>
Very Poor	<div></div>	<div></div>	<div></div>	<div></div>

8. Please provide any additional comments below:

Target Detection/Confidence Questions²

Each time a participant detects a target, he/she will be asked to verbally indicate the target type (shelter, SUV, truck, or van) and their level of confidence that the object selected was in the target set.

1. a. Target type:

- SUV
- Van
- Truck
- Shelter

b. How confident are you that the object was in the target set?

- 1 - not at all confident
- 2 - slightly confident
- 3 - moderately confident
- 4 - fairly confident
- 5 - very confident

² This was not a written questionnaire. Participants provided their target identification and confidence ratings orally. Voice recognition software was used to record their responses.

NASA-TLX Instructions and Questionnaire

We are not only interested in assessing your performance but also the experiences you have during the experimental trials. Right now we are going to describe the technique that will be used to examine these experiences. In the most general sense we are examining the "Workload" you experience. Workload is a difficult concept to define *precisely*, but a simple one to understand *generally*. The factors that influence your experience of workload may come from the task itself, your feelings about your own performance, how much effort you put in, or the stress and frustration you feel. Physical components of workload are relatively easy to conceptualize and evaluate. However, mental components of workload may be more difficult to measure.

Since workload is something that is experienced individually by each person, there are no effective "rulers" that can be used to estimate the workload of different activities. One way to find out about workload is to ask people to describe the feelings they experienced. Because workload may be caused by different factors, we would like you to evaluate several of them individually rather than lumping them into a single, global evaluation of overall workload. This set of six rating scales was developed for you to use in evaluating your experiences during different tasks. (Hand scale sheet on top of explanations to participant)

Please read the descriptions of the scales carefully. If you have a question about any of the scales in the table, please ask me about it. It is important that they be clear to you. You may keep the descriptions with you for reference during the experiment.

(Stop here, read detailed subscale explanations while participant reviews the scale sheet/explanations)

After performing each task, you will evaluate it by marking each scale at the point that matches your experience. Each line has two endpoint descriptors that describe the scale. Note that "performance" goes from "good" on the left to "poor" on the right. This order has been confusing for some people. Mark the desired location. Please consider your responses carefully in distinguishing among the task conditions. When rating each task, only reflect on the one you have just completed. Consider each trial in isolation, that is, do not compare it to prior experiences. Also, please consider each scale individually. Although the definitions may be similar for two or more scales, try to distinguish them from each other based on my explanations and the definitions that you may refer to throughout the experiment- even when rating them.

Your ratings will play an important role in the evaluation being conducted, thus, your active participation is essential to the success of this experiment, and is greatly appreciated!

TLX Weights

For each of the following Pairs, Select the one that represents the higher contributor to workload for the task.

Performance or Temporal Demand

NASATLX

For each category, select a value on the bar by clicking where you want on the bar.

Mental Demand	Low High	How Much mental and perceptual activity was required (e.g., thinking, deciding, calculating, remembering, looking, searching, etc.)? Was the task easy or demanding, simple or complex, exacting or forgiving?
Physical Demand	Low High	How much physical activity was required (e.g., pushing, pulling, turning, controlling, activating, etc.)? Was the task easy or demanding, slow or brisk, slack or strenuous, restful or laborious?
Temporal Demand	Low High	How much time pressure did you feel due to the rate or pace at which the tasks or task elements occurred? Was the pace slow and leisurely or rapid and frantic?
Performance	Good Poor	How successful do you think you were in accomplishing the goals of the task set by the experimenter (or yourself)? How satisfied were you with your performance in accomplishing these goals?
Effort	Low High	How hard did you have to work (mentally and physically) to accomplish your level of performance?
Frustration Level	Low High	How insecure, discouraged, irritated, stressed and annoyed versus secure, gratified, content, relaxed and complacent did you feel during the task?

Practice 1

Next

Figure A-1. NASA-Task Load Index computer application used. Participants rated their perception of the workload for each individual trial in the six categories, descriptions of the categories are located to the right of the scale (top image). When finished with ratings the participant completed a pair wise comparison of the categories (bottom image).

APPENDIX B
Candidate Video Display Formats for DVR Implementation

One technology-based method that has been proposed for improving target detection performance is the use of digital video recording (DVR)/playback. Adding a DVR/playback capability when only one or two live videos are being monitored appears to pose a relatively easy display problem. One could simply display the live video(s) above or next to the DVR/playback which could have some predetermined delay (e.g., 10 seconds behind the live video). Display formats become more difficult when a single operator is required to monitor four or more live videos.

Several four-video presentation conditions will be evaluated. Some candidate display concepts are shown below. In the baseline condition (Figure B-1), participants will view the four live videos without any video playback capability. Each of the other conditions will include different implementations of a video playback capability.

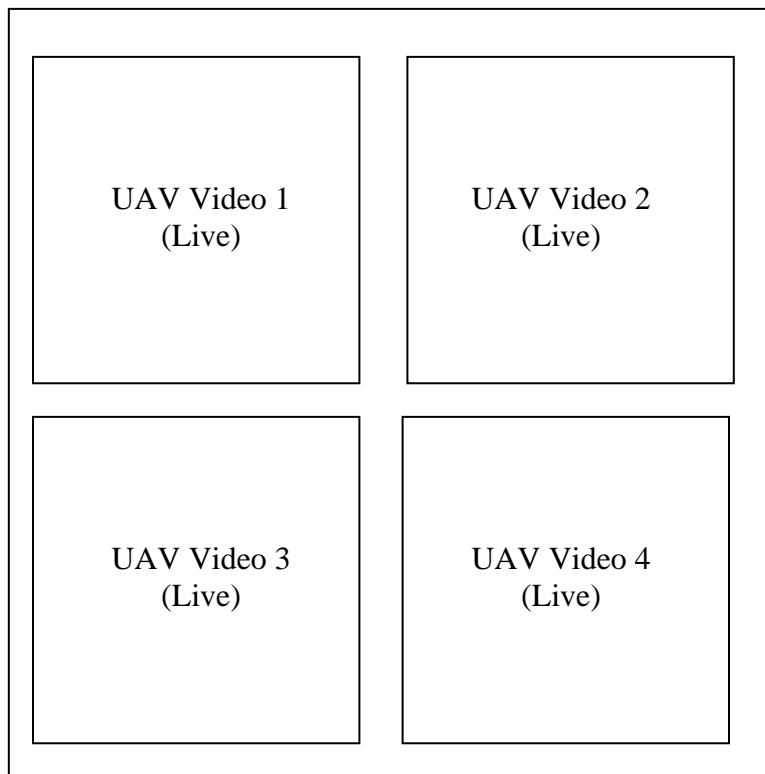


Figure B-1. Four live only video condition (Baseline). Each window = 640 x 480 (pixels).

Figure B-2 shows the first experimental configuration in which the video playback capability is embedded in the same screen as the live video for all four video screens. In this condition, if participants decide to use the video playback capability for a particular video, they will not be able to simultaneously view the live and playback video for that video stream/source. Once they have reviewed the playback video, they can “catch up” to the live portion of the video by using a slider bar to advance the video to the “current” segment of the video.

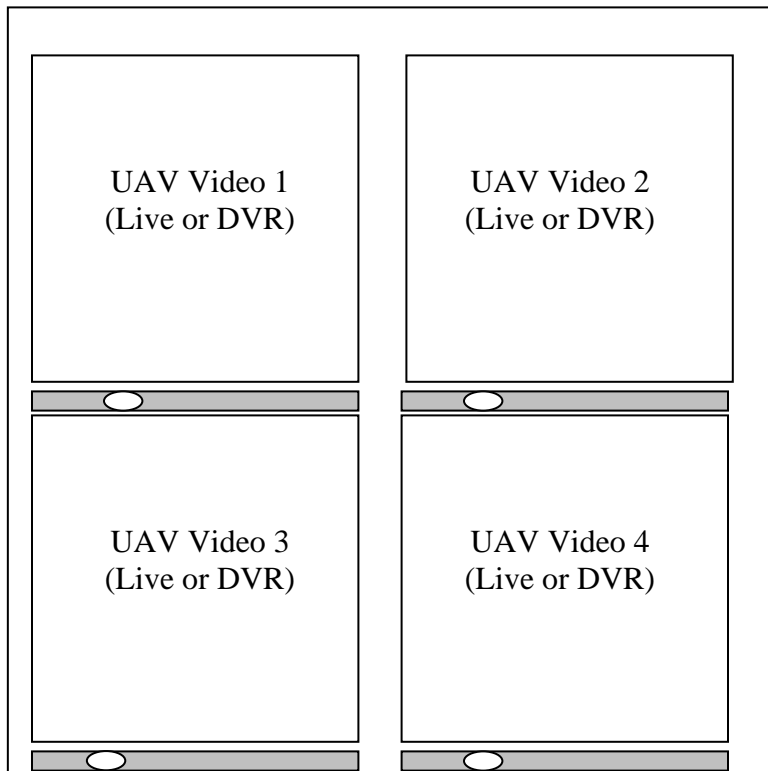


Figure B-2. Four live videos with embedded DVR condition. Each window = 640 x 480 pixels.

In the second experimental configuration, four live videos will be displayed with an additional screen for a single video playback capability. To view the playback for a live video, participants will need to “select” that video source by clicking the mouse on a button below the live feed (Figure B-3).

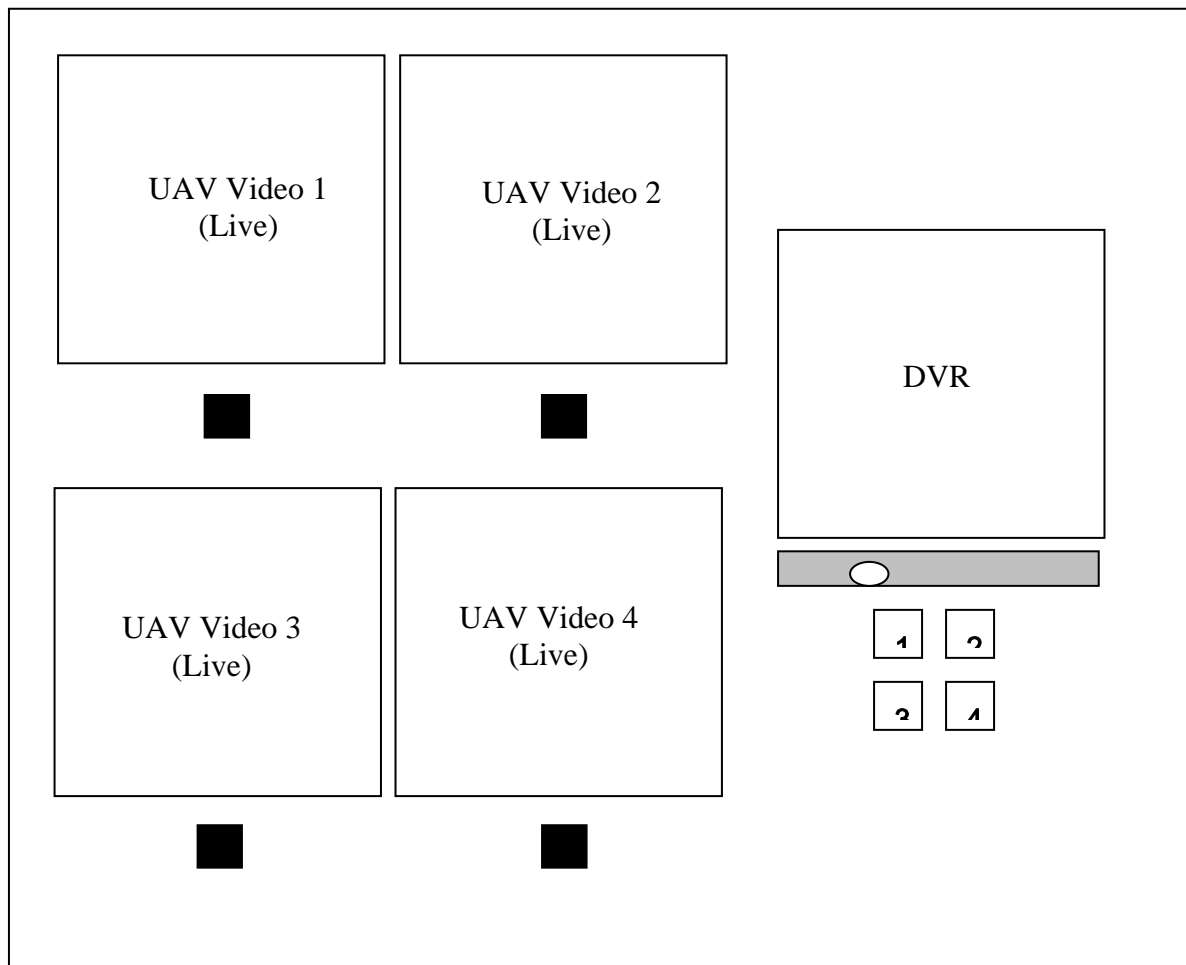


Figure B-3. Four live videos with a single auxiliary DVR capability. Each window = 640 x 480 pixels.

In the third experimental condition, four small windows are presented in the center of the display that show live, reduced-size videos. In addition, four large windows provide a live/DVR capability (Figure B-4).

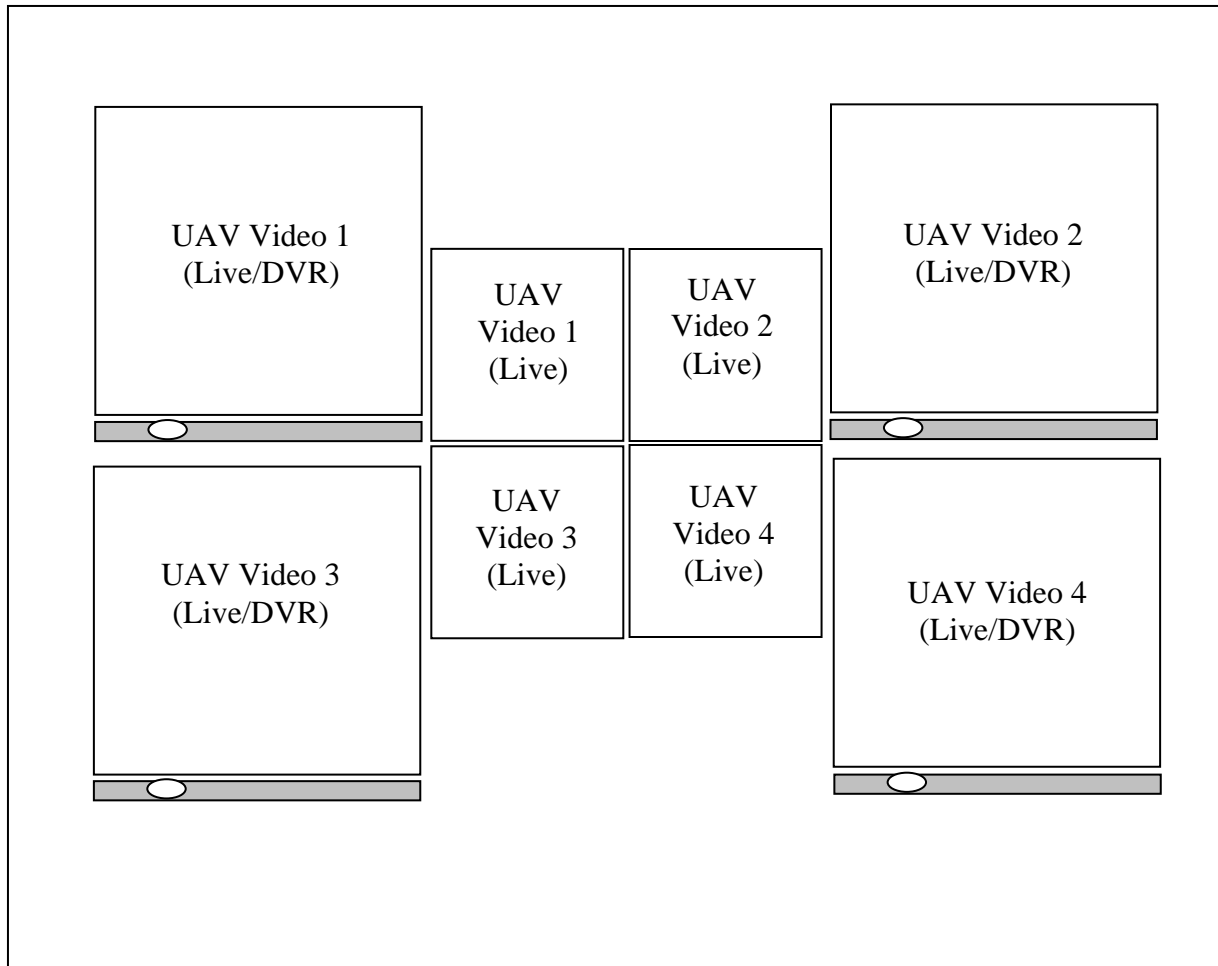


Figure B-4. Four live videos with four auxiliary DVR capabilities condition. Live windows = 315 x 240 (pixels); DVR windows = 640 x 480 (pixels)

In the fourth experimental condition, four live videos are displayed in the upper row of the screen. Corresponding DVRs are shown directly below the live videos. This configuration cannot be implemented on a single screen. It would require two side-by-side screens. It has the advantage that full size displays are used for both the live and DVR videos and that it does not require the operator to perform any actions to select a video for playback. All DVR videos could have a preset delay (e.g., 10 seconds) with the option to change that value to the operator's preferences.

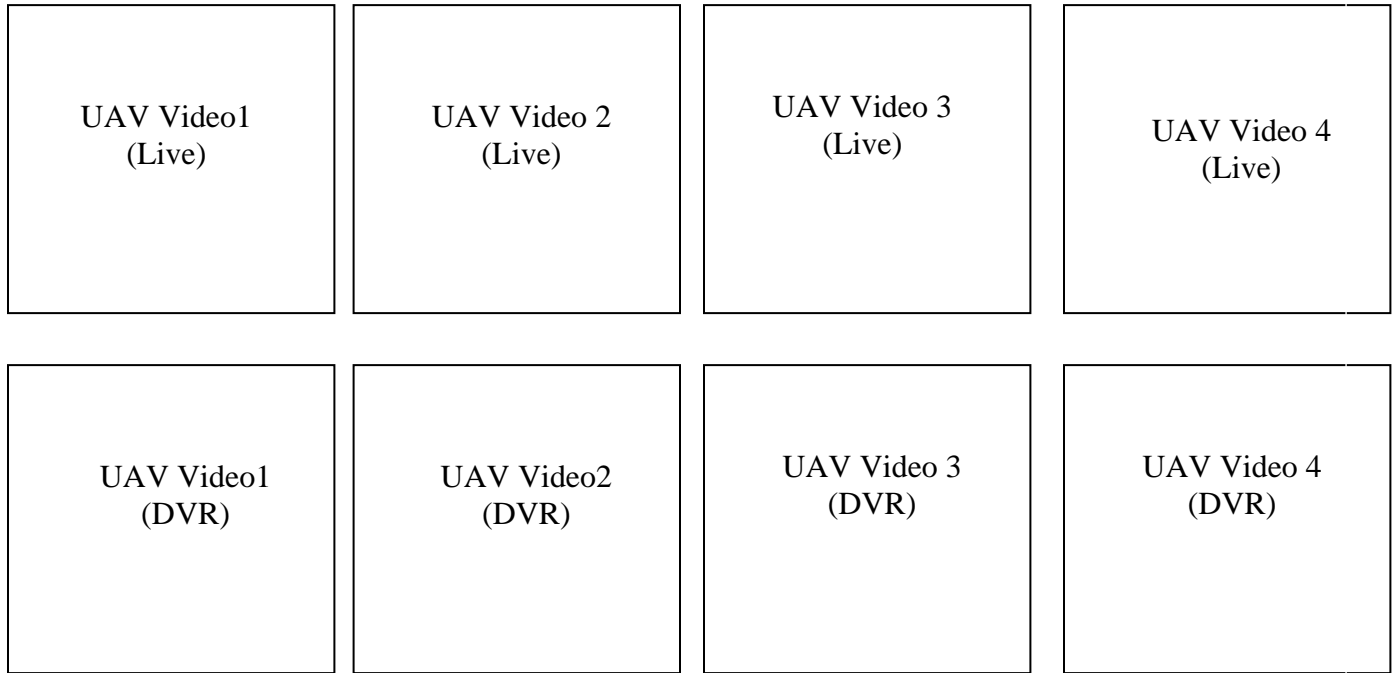


Figure B-5. Four live videos (top) with four auxiliary DVRs capabilities condition. Each window = 640 x 480 (pixels).